

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

USDC SDNY
DOCUMENT
ELECTRONICALLY FILED
DOC #:
DATE FILED: 10/13/2020

CHAD LINDSEY MOSHELL, Individually
and On Behalf of All Others Similarly
Situating,

Plaintiff,

-v-

SASOL LIMITED, DAVID EDWARD
CONSTABLE, BONGANI NQWABABA,
STEPHEN CORNELL, PAUL VICTOR,
and STEPHAN SCHOEMAN,

Defendants.

Case No. 1:20-cv-01008-JPC

Hon. John Peter Cronan

**PROTOCOL GOVERNING THE PRODUCTION OF
ELECTRONICALLY STORED INFORMATION (“ESI”)**

1. **Purpose and Scope.** This Order governs the collection and production of documents (hard copy or electronic) and ESI by the Parties in this Action. Nothing in this Order is intended to be an exhaustive list of discovery obligations or rights of any Party or Non-Party. To the extent additional obligations or rights not addressed in this Order arise under the Federal Rules of Civil Procedure, local rules, applicable state and federal statutes, or subsequent Orders in this litigation, those obligations or rights will control.
2. **Definitions.** The following definitions shall apply to this Protocol:
 1. “Document” is defined consistent with Fed. R. Civ. P. 34 and Local Civil Rule 26.3 (c)(2), and includes ESI and hard copy/paper material. A draft or non-identical copy is a separate Document within the meaning of this term.
 2. “ESI” means electronically stored information, and is defined consistent with Fed. R. Civ. P. 34(a).
 3. “Extracted Text” means the text extracted from a Native File, and includes all header, footer, and document body information.

4. “Load File” means a load utilization file, which is an electronic file containing information identifying a set of paper-scanned images or processed ESI, and containing: (i) an indication of which individual pages or files constitute each Document, including attachments, and links to the Static Images associated with each Document; (ii) links to any Native Files, where native files are being produced, including attachments, associated with each Document; and (iii) data relevant to each individual Document, including extracted and user-created Metadata and coded data.
 5. “Metadata” means: (i) information associated with or embedded in a Native File that does not constitute the primary content region of the file; and (ii) information generated automatically by the operation of a computer or other information technology system when a Native File is created, modified, transmitted, deleted, or otherwise manipulated by a user of such system.
 6. “Native File” or “Native Format” refers to ESI that is produced in the format in which it was maintained (e.g., an Excel document in .xls format would be produced in .xls format).
 7. “OCR” means the optical character recognition file that is created by software used in conjunction with a scanner that is capable of reading text-based documents, and making such documents searchable using appropriate software.
 8. “Producing Party” means the party producing Documents in response to any request for production of documents pursuant to Fed. R. Civ. P. 34(a) or for any other reason.
 9. “Receiving Party” means the party receiving production of Documents in response to any request for production of document(s) pursuant to Fed. R. Civ. P. 34(a) or for any other reason.
 10. “Static Image(s)” means a representation of ESI produced by converting a Native File into a standard image format capable of being viewed and printed on standard document review systems. Tagged Image File Format (TIFF) and Portable Document Format (PDF) images are examples of Static Images.
3. **File Structure.** The Producing Party shall produce the following sets of files with each production:
- a. Index File.
 - (1) Each production has one index file, in “Concordance” style .DAT format.
 - (2) Values must be enclosed by þ (ASCII Decimal 254).
 - (3) Values must be separated by the “Device Control 4” character, ASCII decimal 20.

- (4) First line must contain the column/field names (set forth in Section IV, below).
- (5) The fields Begin Bates and End Bates must be present. The field NativePath must be present when native files are included in the document production.
- (6) Each subsequent row must contain the Metadata for one Document.
- (7) Every row must have the same number of columns/fields (empty values are acceptable).
- (8) Text must be encoded in UTF-16.
- (9) File should be placed in the root directory or a directory labeled "DATA."

b. OCR and Extracted Text Files (.TXT Files).

- (1) A single text file for each Document containing all the Document's pages, in text.
- (2) Pages separated by form feed character (decimal 12, hex 0xC).
- (3) Filenames should be of the form: <Bates num>.txt, where <Bates num> is the Bates number of the first page of the Document.
- (4) Text must be encoded in UTF-16.

Files should be placed in a directory labeled "TEXT."

c. Image Files.

- (1) Single-page Group IV TIFF images for each Document, containing all images for that document.
- (2) Filenames should be of the form: <Bates num>.<ext>, where <Bates num> is the BATES number of the first page of the document (i.e., the "Begin Bates" number), and <ext> is the appropriate extension for the image format (.tiff).
- (3) Files should be placed in the "IMAGES" directory.

4. **Variations.** If any Party identifies a circumstance where application of this Order is not technologically feasible, the Party will disclose to all other Parties the reason(s) for, and circumstances surrounding, the need to vary from this Order, and the Parties will meet and confer in an effort to reach agreement on an appropriate deviation from this Order. In the

event the Parties cannot reach agreement, the matter may be submitted to the Court for determination.

5. **Custodians.** The Parties shall meet and confer in an effort to agree upon the following:
(a) the identity and role of custodians from which documents and ESI will be obtained for production; and (b) the identity, scope and format of custodial and non-custodial sources from which documents and ESI will be obtained for production. The parties shall promptly meet and confer to allow the Parties to identify whether a ruling by the Court is needed on any of the issues (a)-(f) above.
6. **Key Word Searching:** If a Party intends to use “key word” searching of ESI to identify ESI for production, it must notify the Parties of such intent and the Parties will confer regarding search methodologies, search terms, date restrictions, and custodian restrictions. The Parties shall participate in an iterative and cooperative approach in which the Parties will meet and confer regarding reasonable and appropriate methods to increase the relative precision or proportion of relevant and responsive documents within the search results and production sets, including conferring about exchanging “hit count” reports, with the goal of avoiding placing an unnecessary burden on the Producing Party.
7. **File Types.** File types containing text data, such as email stores and individual messages, word processing documents, spreadsheets, hypertext documents and data saved in portable document format (“PDF”) with extractable text will be processed and their text shall be extracted prior to searching. Image files and PDFs containing images of potentially relevant text shall be submitted to Optical Character Recognition (“OCR”) processing before search terms are applied. Documents containing foreign language text must be OCR’d using the

appropriate settings for that language (e.g., OCR of German documents must use settings that properly capture umlauts).

8. **Format.** Documents shall be produced in 300 DPI Group IV Black & White Tagged Image File Format (.TIFF or .TIF) files (“TIFF files”). TIFF files shall be produced in single-page format along with corresponding image load files (.OPT file). For electronically stored information (“ESI”), the TIFF files shall be, where possible, created directly from the original native documents. If an original document contains color, the document should be produced in color as single-page, 300 DPI JPG images with JPG compression and a high quality setting as to not degrade the original image. Parties are under no obligation to enhance an image beyond how it was kept in the usual course of business. The Producing Party may not create TIFF files of electronic documents by printing out paper copies of the electronic documents and scanning those paper copies. For paper documents, the TIFF files shall be created by scanning either the original paper documents or first-generation photocopies of the original paper documents and properly unitizing documents before production. Embedded files shall be produced as a child to the parent document to which it is embedded to the extent responsive to the parties’ requests for documents (“Requests”) and not protected by a claim of privilege. In-line images and OLE objects such as signature blocks from emails shall not be extracted as standalone documents. Notwithstanding the foregoing, all electronically stored spreadsheets, presentation files (e.g. PowerPoint), and multimedia files shall be produced in native electronic format with embedded data and metadata intact and a TIFF placeholder, to the extent responsive to the parties’ Requests and not protected by a claim of privilege. The Parties agree to meet and confer in good faith with respect to the production of data, if any, in the scheduling, cost, contract

management and accounting software platforms used for the Lake Charles Chemical Project (LCCP) megaproject (e.g., Primavera aka “P6” or Sasol’s Contract Management System aka CMSi). Any documents that cannot be converted to TIFF format shall be represented in the production with a placeholder TIFF image which bears the legend “This document cannot be converted to TIFF.” The parties agree to meet and confer regarding such documents if requested by the Receiving Party and to take reasonable actions to remedy such conversion problems.

9. **Request for Native Format.** The Receiving Party may request that the Producing Party provide specific documents or categories of documents in native format after the production of the document in TIFF format. Upon receipt of the request, the Producing Party shall produce the documents in native format, provided that the request is reasonably directed toward the production of a limited number of documents in a usable format. In the event that the Receiving Party requests that a significant volume of documents be produced in native format, the parties shall meet and confer concerning the request for native documents.
10. **Extracted Text.** For all documents, the Producing Party shall provide full extracted text. Where extracted text is unavailable, such as image files and non-searchable PDFs, OCR text shall be provided. Such text files shall be produced as document-level text files and be named consistently with their corresponding TIFF files.
11. **Production Media.** The Producing Party will endeavor to produce documents electronically by way of a secure FTP. To the extent a production by a secure FTP is not possible, the Producing Party shall produce documents on a readily accessible computer or electronic media, including without limitation CD-ROM, DVD, external hard drive

(with standard PC-compatible interface), or such other media as the parties may agree on (“Production Media”). The Producing Party shall affix a unique identifying label to each piece of Production Media, which shall identify the date of the production and the numbered sequence of the material in that production. The Producing Party shall properly package all Production Media to ensure safe shipping and handling. The Producing Party shall encrypt all Production Media prior to shipping.

12. **De-Duplication.** The Producing Party shall de-duplicate identical files based on hash values, at a family level, within and among custodians, using commercially acceptable methods of de-duplication (e.g., MD5 or SHA-1 hash values). If there is any handwriting or other alteration to a document, it shall not be considered a duplicate pursuant to this Paragraph. If the Producing Party becomes aware of any file that was incorrectly filtered during the de-duplication process, the Producing Party shall promptly notify the Requesting Party and produce the file pursuant to the terms of this Agreement. Removal of documents from production using near-deduplication or e-mail thread suppression is not acceptable. Family groups, e.g., an e-mail and its attachments, must be de-duplicated only against other family groups as entities, e.g., using hash values calculated on concatenations of the hash values of all family members, and no document which is not part of a family group shall be de-duplicated against a member of a family group. To the extent documents are de-duplicated and there are duplicate custodians or file paths, the duplicate custodians will be identified in a metadata field titled “Duplicate Custodian(s)” and the duplicate file paths will be identified in a metadata field titled “Duplicate File Path(s).”
13. **Time Zone.** All documents shall be processed in the UTC time zone setting. All metadata pertaining to dates and times will be standard to UTC. The Parties understand and

acknowledge that such standardization affects only dynamic date fields and metadata values and does not affect, among other things, dates and times that are hard-coded text within a file.

14. **Translations.** The parties will meet and confer regarding the translation and production of any documents containing text in a foreign language, including the use of machine translations. A Producing Party must produce any existing English translations of non-English documents that are responsive to the search protocol developed in accordance with the above-captioned action.
15. **Technology Assisted Review.** Predictive coding/technology-assisted review shall not be used for the purpose of culling the documents to be reviewed or produced without notifying the requesting party prior to use and with ample time to meet and confer in good faith regarding a mutually agreeable protocol for the use of such technologies.
16. **Document Numbering.** Each page of each document shall contain a unique prefix which identifies the Producing Party and shall be Bates numbered. For documents produced in TIFF format, the Producing Party shall electronically “burn” a legible, unique Bates number onto each page at a location that does not obliterate, conceal or interfere with any information from the source document. For documents that the Producing Party produces in native format, the beginning Bates number of the TIFF images of those documents shall serve as the file name of the native file. For example, an Excel document produced natively with bates range ABC0001-0005 shall be produced as ABC0001.xls.
17. **Claims of Confidentiality.** All documents shall be produced subject to the Confidentiality Stipulation and Protective Order to be agreed to by the parties and, ultimately, ordered by the Court in the Action (the “Protective Order”). For documents that the Producing Party

produces in TIFF format, if the Producing Party is producing documents subject to a claim that it is protected from disclosure under the Protective Order, the Producing Party shall electronically “burn” the appropriate confidentiality designation onto each page of the document and provide the designation in the metadata load file. If there is a conflict between the provisions of this Agreement and the Protective Order, the Protective Order shall control.

18. **Metadata.** During the process of converting electronic documents from the electronic format of the application in which the document is created, viewed and/or modified to TIFF, metadata values shall be extracted and produced in a metadata load file (.DAT file using concordance standard delimiters). The metadata load file shall contain a link to natively produced documents via data values called “Native Link.” The Native Link values should contain the full directory path and file name of the documents as contained in the produced media. The Native Link field should be included in the .DAT file. The following metadata fields shall be produced if available or where practicable through each party's usual method of data extraction and if they do not reveal privileged information:

Field Name	Description
BatesBegin	The first page of the document
BatesEnd	The last page of the document
AttachBegin	The first page/doc of the first parent of an attachment family (e.g., ABC000001)
AttachEnd	The last page/doc of the last attachment (e.g., ABC000019)
AttachCount	Number of document attachments.
Pages	Total number of pages in the document.
Placeholder	Identifies a document has a placeholder image (y/n).
Custodian	Custodian name

Field Name	Description
Duplicate Custodian(s)	Duplicate custodians of a globally de-duplicated document separated by semicolons (e.g., Doe, John; Roe, Robert)
Tag – Confidentiality	The text demonstrating the confidential level of the document (e.g., Confidential, Highly Confidential, etc.)
Time_Zone	The time zone in which the emails were standardized during conversion
Author	Author field extracted from the metadata of a Document or other creator identified for the Document.
EM – Subject	E-mail subject
EM – To	The values in the original “To” field for emails
EM – From	The values in the original “From” field for emails
EM – CC	The values in the original “CC” field for emails
EM – BCC	The values in the original “BCC” field for emails
EM – Time Sent	Time the e-mail was sent in UTC.
EM – Date Sent	Date the email was sent
EM – Date Received	Date the email was received
EM – Time Received	Time the email was received in UTC.
File – Title	Title field value extracted from the metadata of the native file
File – Document Type	A reference to the application that created the file, for example “Word Document” or “Excel Spreadsheet”
File – File Name	Original file name
File – File Path	The path of the original native file (excluding name and file extension)
File – Duplicate File Path	The path of the original duplicate native file (excluding name and file extension)
File – File Size	File size of the native file in bytes
File – Date Created	Date the file was created (not applicable to emails that were sent or received)
File – Date Modified	Date the file was last modified (not applicable to emails that were sent or received)
Last Modified By	Identification of person(s) who last modified a document.
File Extension	File extension of the document (e.g., xls)

Field Name	Description
Extracted TextPath	File path to the extracted text/OCR file, or the extracted/OCR file link (e.g., ABC001\Text\001\ABC000001.txt)
Hidden Fields	Shows if there are: <ul style="list-style-type: none"> • Excel Comments • Excel Hidden Rows • Excel Hidden Worksheets • Excel Very Hidden Worksheets • PDF Crop Box • Power Point Hidden Slides • Power Point Speaker Notes • Word Comments • Word Hidden Text • Word Revisions
Native File Path	File path to the native file, or the native file link (e.g., ABC001\Native\001\ABC000001.xls)
MD5Hash	Unique “fingerprint” that exists for every document and will be used for identification of exact duplicate documents
SHA1 Hash	The SHA1 hash value of a document
Is Embedded	Yes/No indicator of whether a file is embedded in another document
Redacted	If a document contains a redaction, this field will display ‘Yes’
Production Volume	Production volume number (e.g., V001, V002, etc.).
Producing Party	Name of party producing the document.
Track Changes	Document has track changes (Y/N).
Has Comments	Indicates there are comments in the document.

All date fields will be formatted MM/DD/YYYY and all time fields will be formatted HH:MM:SS and converted to UTC. To the extent reasonably available, the “Custodian” or “File Path” field with respect to ESI gathered from an individual’s hard drive will provide metadata sufficient to identify the individual custodian or non-custodial source from whose hard drive such ESI has been gathered.

19. **Native File Production.** For documents produced in their native file format, a placeholder TIFF image shall be produced for these records bearing the legend “Produced in Native File Format Only.” The TIFF image shall be endorsed with a sequential Bates number and the produced native file named to match this Bates number.
20. **Decryption and Passwords.** The Producing Party may employ reasonably available electronic measures to break the encryption of, or unlock, any encrypted or password protected document and may clean any document infected by a virus before making a production of such a document, notwithstanding the fact that such measures may alter metadata fields listed in Paragraph 11 and alter the native format of the document.
21. **Privilege Log.** The parties will provide a privilege log for any documents withheld in full based upon a claim of privilege in accordance with Local Civil Rule 26.2, including that “[e]fficient means of providing information regarding claims of privilege are encouraged . . . [f]or example, when asserting privilege on the same basis with respect to multiple documents, it is presumptively proper to provide the information required by this rule by group or category. A party receiving a privilege log that groups documents or otherwise departs from a document-by-document or communication-by-communication listing may not object solely on that basis, but may object if the substantive information required by this rule has not been provided in a comprehensible form.”
22. **Redacted Documents.** Documents may only be redacted on the basis of attorney-client privilege, work-product doctrine, or other legally recognized privilege, protection, or immunity. Documents redacted on the basis of privilege shall not be listed on the privilege log. Instead, the parties will apply redactions so as to preserve as much context as possible so that the information traditionally available from a privilege log will appear on the face

of the document, including the sender, recipient, the date of the document, and a stamp designating the basis of the privilege assertion, including: (a) AC for Attorney/Client, (b) WP for Attorney Work Product, (c) CI for Common Interest; and/or (d) OT for Other. Moreover, to the extent the metadata fields are not privileged, all redacted documents will be produced with standard metadata fields intact, preserving all custodial data, document dates, and file names. The Parties shall also produce an auto-generated chart for redacted documents, which will contain the following fields: BatesBegin, BatesEnd, EM – To, EM – From, EM – CC, EM – BCC, EM – Date Sent, and File – Document Type. No party shall make any redactions based on unresponsiveness, except to the extent necessary to preserve Personally Identifying Information (“PII”), sensitive health information, or other similar information. If the Producing Party seeks to make redactions based on unresponsiveness for any other reason, the Receiving Party agrees to consider the request in good faith and meet and confer with the Producing Party concerning the request.

SO STIPULATED AND AGREED.

Dated: October 8, 2020

/s/ Steve. W. Berman

Steve W. Berman

HAGENS BERMAN SOBOL SHAPIRO LLP

1301 Second Avenue, Suite 2000

Seattle, WA 98101

Telephone: (206) 623-7292

steve@hbsslaw.com

*Counsel for Lead Plaintiff David Cohn and
Additional Plaintiff Representative Chad
Lindsey Moshell*

/s/ Caroline H. Zalka

Caroline H. Zalka

WEIL, GOTSHAL & MANGES LLP

767 Fifth Avenue

New York, New York 10153

Tel: (212) 310-8000

caroline.zalka@weil.com

*Counsel for Defendants Sasol Limited,
David Edward Constable, Bongani
Nqwababa, Stephen Cornell, Paul Victor,
and Stephan Schoeman*

SO ORDERED.

Dated: New York, New York

October 13, 2020



John Peter Cronan, U.S.D.J.